ELSEVIER Theoretical Computer Science 142 (1995) 257–276

Theoretical Computer Science

Some undecidable termination problems for semi-Thue systems☆

Géraud Sénizergues*

LaBRI, Université de Bordeaux I, 351 Cours de la Libération, 33405 Talence, Cedex, France

Abstract

We show that the uniform termination problem is undecidable for length-preserving semi-Thue systems having 9 rules. We then give an explicit uniformly terminating semi-Thue system $\mathscr{T}$ having 9 rules which is "universal with respect to termination problems" in some sense. It follows that there exists a fixed rule (u_0, v_0) such that $\mathscr{T} \cup \{(u_0, v_0)\}$ has 10 rules and undecidable termination problem.

1. Introduction

It is known that the *uniform termination* problem for finite semi-Thue systems is undecidable ([16, point (8) p. 227] or [18]) and that there exists some fixed finite semi-Thue system with undecidable *termination* problem (follows easily from the undecidability of the "*halting* problem" for Turing machines [13, Theorem 2.2, p. 70] and the translation of Turing machines into semi-Thue systems [13, Section 2, p. 88–93]; see [18]). These general undecidability results have been recently refined in two directions:

1. The uniform termination problem for *length-preserving* semi-Thue systems is undecidable [8]; this result implies the undecidability of the uniform termination problem for the finite graph rewriting systems studied in [1].

2. There exists a *term* rewriting system with only *one rule* and undecidable termination problem [12].

Our work will follow this trend towards a better knowlede of the frontier between decidable and undecidable problems. We establish here that:

3. The uniform termination problem is undecidable for length-preserving finite semi-Thue systems having 9 rules (Theorem 9).

☆ This work has been supported by the ESPRIT Basic Research Working Group "COMPUGRAPH II".
* Email: ges@labri.u-bordeaux.fr.

SSDI 0304-3975(94)00278-9

4. There exists a semi-Thue system having 10 rules and with undecidable termination problem (Theorem 19).

The proof of (3) consists in giving a reduction from the *accessibility* problem for semi-Thue systems to the *uniform termination* problem for length-preserving semi-Thue systems; this reduction shows that, if there exists some semi-Thue system with n rules and undecidable accessibility problem, then the u-termination problem is undecidable for length-preserving semi-Thue systems with $n+4$ rules (Theorem 8). The proof given by Matiyasevich that there exists some Thue system $\mathcal{M}$ with 3 rules and undecidable *word* problem [21, 10, 5, 22] shows as well that there exists some semi-Thue system $\mathcal{M}$ with 5 rules and undecidable accessibility problem (this was noticed in [23]). Our result (3) then follows.[1]

Result (4) is obtained by an adaptation of the ideas of [21]. We give an explicit construction of a uniformly terminating system $\mathcal{T}$, which has 9 rules and is "universal with respect to termination problems" in the following sense: to every finite semi-Thue system S and word t we can effectively associate a rule $u_S \to v_S$ and a word $w_{S,t}$ such that S has an infinite derivation starting from t iff $\mathcal{T} \cup \{(u_S \to v_S)\}$ has an infinite derivation starting from $w_{S,t}$ (Theorem 18). Our result (4) follows easily.

To our knowledge, it is still unknown whether each 1-rule semi-Thue system has a decidable termination problem (question raised in [12, p. 110]), hence the frontier between decidability and undecidability (in this area) lies between 0 and 10. As well, it is unknown whether the u-termination problem is decidable for 1-rule semi-Thue systems (but it is clearly decidable in the *length-preserving* case), hence the frontier between decidability and undecidability (in this area) lies between 0 and 9. In the length-preserving case this frontier lies between 1 and 9.

2. Preliminaries

2.1. Vocabulary, notation

Words: Given an alphabet X, by $(X^*, ., \varepsilon)$ we denote the *free monoid* generated by X (where . is the concatenation product and ε is the empty word). As X^* is embedded in $F(X)$ (the *free group* generated by X) we sometimes use the notation u^{-1} for the *inverse* in $F(X)$ of a word u. If u denotes a word, $\tilde{u}$ denotes its *mirror* image.

Semi-Thue systems: A *semi-Thue* system over X is a subset $S \subset X^* \times X^*$. By $\to_S$ we denote the binary relation $\forall f, g \in X^*$, $f \to_S g$ iff there exist $(u, v) \in S$ and $\alpha, \beta \in X^*$ such that $f = \alpha u \beta$, $g = \alpha v \beta$. $\to_S$ is the *one-step* rewriting relation generated by S. The pair (α, u) is called the *redex* used in the rewriting step $f \to_S g$. $\overset{i}{\to}_S$ (where $i \in \mathbb{N}$), $\overset{*}{\to}_S$, $\overset{+}{\to}_S$ are then defined as usual from $\to_S$ (see [17]). By $\leftrightarrow_S$ we denote the binary relation:

[1] The undecidability bounds given in [6, 11] concerning the word problem for groups and in [23] concerning the Post-correspondence problem are leaning on the result of [21] too.

$\forall f, g \in X^*, f \leftrightarrow_S g$ iff $f \rightarrow_S g$ or $g \rightarrow_S f$. $\overset{i}{\leftrightarrow}_S, \overset{*}{\leftrightarrow}_S, \overset{+}{\leftrightarrow}_S$ are then defined as usual from $\leftrightarrow_S$. A *derivation* (mod S) is a sequence $D = (w_i, \alpha_i, u_i, v_i)_{i \in I}$ where I is a nonempty beginning section of $\mathbb{N}$ (i.e. $I = [0, n]$ for some $n \in \mathbb{N}$ or $I = \mathbb{N}$), for every $i \in I$, $w_i \in X^*$, $(u_i, v_i) \in S$ and $(i + 1 \in I) \Rightarrow (\alpha_i u_i$ is a prefix of w_i, $w_{i+1} = \alpha_i v_i (\alpha_i u_i)^{-1} w_i)$. The *length* of D is n (if $I = [0, n]$) or ∞ (if $I = \mathbb{N}$). D is said to be a *rl-derivation* (rl stands for right-to-left) iff, $\forall i \in I - \{0\}$, $|\alpha_i| < |\alpha_{i-1} v_{i-1}|$. In order to abbreviate the notation we often (incorrectly) drop the data α_i, u_i, v_i in the definition of a derivation D. A semi-Thue system is called *homogeneous*[2] iff, $\forall (u, v), (u', v') \in S$, $|u| = |u'|$ and $|v| = |v'|$.

Algorithmic problems: The following algorithmic problems on semi-Thue systems are classical.

(TP) *Termination problem*: The termination problem for the alphabet X and the semi-Thue system $S \subset X^* \times X^*$ is the following:

Instance: $w \in X^*$.
Question: Does every derivation (mod S) starting on w have finite length?

(when the answer is "yes", we say that S terminates on w).

(UTP) *Uniform termination problem*: The uniform termination problem for a class $\mathscr{C}$ of semi-Thue systems is the following:

Instance: An alphabet X and a finite semi-Thue system $S \subset X^* \times X^*$ which belongs to $\mathscr{C}$.
Question: Does every derivation (mod S) starting from a word in X^* have finite length?

(When the answer is "yes" we say that S is *uniformly terminating*, some-times abbreviated as u-terminating.)

(NUTP) *Nonuniform termination problem*: The nonuniform termination problem for a class $\mathscr{C}$ of semi-Thue systems is the following:

Instance: An alphabet X and a finite semi-Thue system $S \subset X^* \times X^*$ which belongs to $\mathscr{C}$.
Question: Does there exist some infinite derivation (mod S) starting from a word in X^*?

(WP) *Word problem*: The word problem for the alphabet X and the semi-Thue system $S \subset X^* \times X^*$ is the following:

Instance: Two words $w_1, w_2 \in X^*$.
Question: $w_1 \overset{*}{\leftrightarrow}_S w_2$?

[2]This terminology extends the terminology used in [2, 3] for example.

(IWP) *Individual word problem*: The individual word problem for the alphabet X, the semi-Thue system $S \subset X^* \times X^*$ and the word $w \in X^*$ is the following:

Instance: One word $w_1 \in X^*$.
Question: $w_1 \overset{*}{\leftrightarrow}_S w$?

(ACP) *Accessibility problem*: The accessibility problem for the alphabet X and the semi-Thue system $S \subset X^* \times X^*$ is the following:

Instance: Two words $w_1, w_2 \in X^*$.
Question: $w_1 \overset{*}{\rightarrow}_S w_2$?

(IAP) *Individual accessibility problem*: The individual accessibility problem for the alphabet X, the semi-Thue system $S \subset X^* \times X^*$ and the word $w \in X^*$ is the following:

Instance: One word $w_1 \in X^*$.
Question: $w_1 \overset{*}{\rightarrow}_S w$?

Let us notice that the WP for S (resp. the IWP for S, w) trivially reduces to the ACP for $S' = \{(u,v) \in X^* \mid (u,v) \in S \text{ or } (v,u) \in S\}$ (resp. the IAP for S', w). For more information the reader should refer to [19,4] (about semi-Thue systems) or [14,15] (about termination problems).

2.2. Some useful results and systems

Lemma 1 (rl-derivation lemma). *Let S be a finite subset of $X^+ \times X^*$ and let $D = (w_i, \alpha_i, u_i, v_i)_{i \in I}$ be a derivation* (mod S).

1. *If I is finite, there exist some permutation $\sigma : I \to I$ and some sequence $(w'_i, \alpha'_i)_{i \in I}$ such that $w'_0 = w_0$ and $(w'_i, \alpha'_i, u_{\sigma(i)}, v_{\sigma(i)})_{i \in I}$ is a rl-derivation* (mod S)
2. *If $I = N$, there exist some injection $\sigma : I \to I$ and some sequence $(w'_i, \alpha'_i)_{i \in I}$ such that $w'_0 = w_0$ and $(w'_i, \alpha'_i, u_{\sigma(i)}, v_{\sigma(i)})_{i \in I}$ is a rl-derivation* (mod S).

Proof. (1) *Finite derivations*: Let us define an equivalence relation $\sim$ on the set of finite derivations (mod S) by:

For every $D = (w_i, \alpha_i, u_i, v_i)_{i \in I}$, $D' = (w'_j, \alpha'_j, u'_j, v'_j)_{j \in J}$, $D \sim D'$ iff $I = J = [0,n]$ (for some $n \geqslant 0$), $w_0 = w'_0$, $w_n = w'_n$ and there exists some permutation $\sigma \in \mathscr{S}_I$ such that

$$\forall i \in I, \quad (u_{\sigma(i)}, v_{\sigma(i)}) = (u'_i, v'_i).$$

Let us denote by $\preceq_n$ the lexicographic ordering on $\mathbb{N}^n$ induced by the ordering $\geqslant$ on $\mathbb{N}$:

$$\boldsymbol{p} = (p_1, p_2, \ldots, p_n) \preceq_n \boldsymbol{p}' = (p'_1, p'_2, \ldots, p'_n)$$
$$\Leftrightarrow \boldsymbol{p} = \boldsymbol{p}' \text{ or } \exists j \in [1,n] \mid \forall i < j,\ p_i = p_j \text{ and } p_j > p'_j.$$

For every $D = (w_i, \alpha_i, u_i, v_i)_{i\in[0,n]}$, we note

$$\boldsymbol{a}(D) = (|\alpha_0|, |\alpha_1|, \ldots, |\alpha_{n-1}|).$$

Let us show now that

$$D \text{ is not rl} \Rightarrow \exists D' \text{ such that } D' \sim D \text{ and } \boldsymbol{a}(D') \prec_n \boldsymbol{a}(D). \tag{1}$$

If D is not rl, then $n \geqslant 2$ and $\exists k\in[1, n-1] \mid w_{k-1} = \alpha_{k-1}u_{k-1}\gamma_{k-1}u_k s_{k-1}$, $w_k = \alpha_{k-1}v_{k-1}\gamma_{k-1}u_k s_{k-1}$, $w_{k+1} = \alpha_{k-1}v_{k-1}\gamma_{k-1}v_k s_{k-1}$, where $\alpha_{k-1}v_{k-1}\gamma_{k-1} = \alpha_k$.

Let us define

$$w'_i = w_i \ (i \neq k), \quad w'_k = \alpha_{k-1}u_{k-1}\gamma_{k-1}v_k s_{k-1},$$

$$\alpha'_i = \alpha_i \ (i \notin \{k-1, k\}), \quad \alpha'_{k-1} = \alpha_{k-1}u_{k-1}\gamma_{k-1}, \quad \alpha'_k = \alpha_{k-1},$$

$$(u'_i, v'_i) = (u_i, v_i) \ (i \notin \{k-1, k\}), \quad (u'_{k-1}, v'_{k-1}) = (u_k, v_k), \quad (u'_k, v'_k) = (u_{k-1}, v_{k-1}).$$

Then $D' = (w'_i, \alpha'_i, u'_i, v'_i)_{i\in[0,n]}$ is a derivation (mod S) such that

$$D' \sim D \quad \text{and} \quad \boldsymbol{a}(D') \prec_n \boldsymbol{a}(D)$$

because, $\forall i < k-1, |\alpha_i| = |\alpha'_i|$ and $|\alpha_{k-1}| < |\alpha'_{k-1}|$. Hence our assertion (1) is proved.

Now let D be any derivation of length n. As $(\mathbb{N}^n, \preceq_n)$ is a well-ordered set, there exists some $D' \sim D$ such that

$$\boldsymbol{a}(D') = \min_{\preceq_n}\{\boldsymbol{a}(D'') \mid D'' \sim D\}.$$

By the assertion (1) D' is rl.

(2) *Infinite derivations*: Let $D = (w_i, \alpha_i, u_i, v_i)_{i\in\mathbb{N}}$ be a derivation (mod S). Let E be the set of all the finite rl-derivations D' fulfilling the following:

D' is a finite rl-derivation (mod S) of the form $(w'_i, \alpha'_i, u'_i, v'_i)_{i\in[0,n]}$ such that there exists some injection $\sigma : [0, n] \to \mathbb{N}$ with $\forall i\in[0, n]$, $(u_{\sigma(i)}, v_{\sigma(i)}) = (u'_i, v'_i)$.

Let $\prec$ be the binary relation on E defined by:

For every $D' = (w'_i, \alpha'_i, u'_i, v'_i)_{i\in[0,n]}$, $D'' = (w''_i, \alpha''_i, u''_i, v''_i)_{i\in[0,m]}$, $D' \prec D''$ iff $n+1 = m$ and $\forall i\in[0, n]$, $(u'_i, v'_i) = (u''_i, v''_i)$.

As S is *finite*, the relation $\prec$ is *finitely branching*:

$$\forall D'\in E, \quad \{D''\in E \mid D' \prec D''\} \text{ is a finite set.} \tag{2}$$

The set E endowed with the relation $\prec$, considered as a directed graph, is infinite (by point (1) of the lemma) and finitely branching. Moreover, there are only finitely many derivations $D'\in E$ of length 0. Hence one such derivation D'_0 of length 0 is a root of an infinite subgraph of E and, by König's lemma [20], this subgraph contains an infinite path:

$$D'_0 \prec D'_1 \prec \cdots \prec D'_k \prec \cdots, \tag{3}$$

where $D'_k = (w'_{k,i}, \alpha'_{k,i}, u'_{k,i}, v'_{k,i})_{i\in[0,k]}$. Let us define then

$$\bar{D} = (\bar{w}_i, \bar{\alpha}_i, \bar{u}_i, \bar{v}_i)_{i\in\mathbb{N}},$$

where $\bar{w}_i = w'_{i,i}$, $\bar{\alpha}_i = \alpha'_{i,i}$, $(\bar{u}_i, \bar{v}_i) = (u'_{i,i}, v'_{i,i})$. As the sequence (D'_k) fulfills (3), D is a derivation (mod S). Let us denote by $\sigma_k : [0,k] \to \mathbb{N}$ an injection such that $\forall j \in [0,k]$, $(u'_{k,j}, v'_{k,j}) = (u_{\sigma_k(j)}, v_{\sigma_k(j)})$. We then define inductively a function $\sigma : \mathbb{N} \to \mathbb{N}$ by

$$\sigma(0) = \sigma_0(0), \tag{4}$$

$$\forall i \geqslant 1, \quad \sigma(i) = \min\{s \in \sigma_i([0,i]) - \sigma([0,i-1]) \mid (u_s, v_s) = (\bar{u}_i, \bar{v}_i)\}. \tag{5}$$

We have to show that this definition is sound, i.e. the set used on the right-hand side of definition (5) is not empty.

As $\bar{D}$ and D'_i are equal up to the index $i-1$,

$$\{j \in [0,i-1] \mid (u_{\sigma_i(j)}, v_{\sigma_i(j)}) = (\bar{u}_i, \bar{v}_i)\} = \{j \in [0,i-1] \mid (u_{\sigma(j)}, v_{\sigma(j)}) = (\bar{u}_i, \bar{v}_i)\}.$$

But the restrictions of σ_i, σ to $[0,i-1]$ are both injective, hence

$$\begin{aligned}&\mathrm{Card}\{s \in \sigma_i([0,i-1]) \mid (u_s, v_s) = (\bar{u}_i, \bar{v}_i)\}\\ &\quad = \mathrm{Card}\{s \in \sigma([0,i-1]) \mid (u_s, v_s) = (\bar{u}_i, \bar{v}_i)\}.\end{aligned}$$

In addition, $(u_{\sigma_i(i)}, v_{\sigma_i(i)}) = (\bar{u}_i, \bar{v}_i)$, hence,

$$\{s \in \sigma_i([0,i]) - \sigma([0,i-1]) \mid (u_s, v_s) = (\bar{u}_i, \bar{v}_i)\} \neq \emptyset.$$

By definition (5), σ is injective. Hence we have proved that, under the assumption that there exists some σ defined on $[0,i-1]$, injective, fulfilling Eqs. (4) and (5) for all $j \leqslant i-1$, σ can be extended to $[0,i]$ so as to be injective, fulfilling Eqs. (4) and (5) for all $j \leqslant i$. The existence of an injection σ defined on $\mathbb{N}$ and fulfilling (4) and (5) follows. The sequence $(\bar{w}_i, \bar{\alpha}_i)_{i \in \mathbb{N}}$ and the injection σ are fulfilling the conclusion of the lemma (point (2)). □

Remarks. (1) The conclusions of Lemma 1 can fail when the hypothesis "$S \subset X^+ \times X^*$" is no more fulfilled: if $S = \{(\varepsilon, \varepsilon)\}$, $\varepsilon \to_S \varepsilon$, (resp. $\varepsilon \to_S \varepsilon \to_S \cdots$) is a finite (resp. infinite) derivation D such that conclusion (1) (resp. (2)) does not hold. Anyway, if S contains ε as a left-hand side of rule, it is clearly not u-terminating.

(2) In conclusion (2) of the lemma, we cannot strengthen the conditions on σ so as to have a *permutation* σ: let $S = \{(a,b),(b,a),(c,d),(d,c)\}$ on $X = \{a,b,c,d\}$. Then the infinite periodic derivation

$$ac \to_S ad \to_S bd \to_S bc \to_S ac \to_S \cdots$$

is such that there is no *permutation* σ fulfilling conclusion (2).

(3) We believe the conclusions of Lemma 1 remain true even for an *infinite* semi-Thue system S, though our proof of point (2) fails in this case (König's lemma cannot be used any more); one could write such a proof, using the notion of planar directed acyclic graph associated to a derivation (as is done in [7] for context-sensitive grammars) and showing the existence of an ad hoc "right-to-left" infinite path in such an infinite DAG.

Definition (*System* $\mathscr{U}$ [21]). Let $\{a,b,e\}$ be an alphabet and $\mathscr{U}$ be the set of rules:

$$eaa \to ae, \quad eab \to be, \quad eba \to ae, \quad ebb \to be.$$

$\mathscr{U}$ is clearly confluent; we remind the reader it means that:

$\forall u,v,w \in \{a,b,e\}^*$, if $u \xrightarrow{*}_{\mathscr{U}} v$ and $u \xrightarrow{*}_{\mathscr{U}} w$ then

$\exists u' \in \{a,b,e\}^*$ such that $v \xrightarrow{*}_{\mathscr{U}} u'$ and $w \xrightarrow{*}_{\mathscr{U}} u'$.

$\mathscr{U}$ is clearly u-terminating too. We denote by ρ the "total reduction" associated to $\mathscr{U}$:

$\forall w \in \{a,b,e\}^*$, $\rho(w)$ is the unique $\mathscr{U}$-irreducible word such that $w \xrightarrow{*}_{\mathscr{U}} \rho(w)$.

Lemma 2 (Folding lemma [21]). *Let $k \geqslant 0$. Let $(C_j)_{j \in [1,2^k]}$ be words all having the same length n: $C_j = c_{1,j}c_{2,j}\ldots c_{n,j}$ (where $c_{i,j} \in \{a,b\}$).*

Let $N = c_{1,1}c_{1,2}\cdots c_{1,2^k}c_{2,1}c_{2,2}\ldots c_{2,2^k}\ldots c_{n,1}c_{n,2}\ldots c_{n,2^k}$. Let $u,v \in \{a,b\}^*$ such that $|u| + |v| = 2^k - 1$ and $x \in \{a,b\}$. Then $\rho(e^k uNvx) = C_j x e^k$ with $j = |vx|$.

This lemma is implicit in [21]; a very similar statement is proved in [5, Lemma 2.5 p. 35].

Definition (*The encoding* τ). Let X be an alphabet containing $\{a,b,e\}$ and let $\tau : X^* \to X^*$ be the homomorphism defined by

$$\tau(a) = a, \quad \tau(b) = ab, \quad \tau(e) = bb, \quad \forall x \in X - \{a,b,e\}, \ \tau(x) = x.$$

As $\eta_{|X}$ is injective and $\tau(X)$ is a suffix code, τ is an injective homomorphism. Let us note $\theta : \{a,b\} \to \{a,b\}$ the bijection exchanging a and b.

Lemma 3 (Lifting lemma [21]). *Let $S \subset X^* \times X^*$ be a semi-Thue system such that*

(I) $\forall x \in \{a,b\}$, $\forall u,w \in X^*$, $\forall n \geqslant 0$,
if $(uxe^n, w) \in S$ then $\exists v \in X^*$, $\exists m \geqslant 0$ such that
(i) $w = vxe^m$,
(ii) $(u\theta(x)e^n, v\theta(x)e^m) \in S$,

(II) $\forall (u,v) \in S$, $u \notin e^*$.

Then the following is true:

$\forall w_1 \in X^*$, $\forall t \in X^*$, *if* $\tau(w_1) \to^*_{\tau(S)} t$, *then* $\exists w_2 \in X^*$, $t = \tau(w_2)$ *and* $w_1 \xrightarrow{*}_S w_2$.

This lemma is similar to [5, Lemma 2.16, p. 39].

Proof. Let us suppose that $\tau(w_1) = \alpha\tau(u_1)\beta$ where $\alpha,\beta \in X^*$, $(u_1,v_1) \in S$. Hypothesis II implies that there exists $t_1,s_1 \in X^*$, $y \in X - \{b,e\}$, $\tau(u_1) = t_1 y s_1$. As the only occurrence of y in a word of $\tau(X)$ is in first position (from the left of the word), $t_1 \in \mathrm{Im}(\tau)$, $\alpha t_1 \in \mathrm{Im}(\tau)$ and $\tau^{-1}(\alpha t_1)$ is a prefix of w_1. Let us note $u'_1 = \tau^{-1}(\alpha)$.

Case 1: $u_1'u_1$ is a prefix of w_1 (i.e. $\exists u_1'' \in X^*$, $w_1 = u_1'u_1u_1''$). Then $w_2 = u_1'(v_1)u_1''$ has the required property.

Case 2: $u_1'u_1$ is not a prefix of w_1. As u_1' is a prefix of w_1, one of the following subcases occurs:

Subcase 2.1: $u_1 = uae^n$ where $u \in X^*$, $n \geqslant 0$ and $w_1 = u_1'(ube^n)u_1''$, where $u_1'' \in X^*$. By hypothesis I(i), $v_1 = vae^m$ and by I(ii) the word $w_2 = u_1'(vbe^m)u_1''$ has the required property.

Subcase 2.2: $u_1 = ube^n$ where $u \in X^*$, $n \geqslant 0$ and $w_1 = u_1'(uae^n)u_1$ where $u_1'' \in X^*$. Then $v_1 = vbe^n$ and the word $w_2 = u_1'(vae^m)u_1''$ has the required property. □

Definition (*System* $\mathscr{M}$). In [21] a semi-Thue system $\mathscr{M}$ is defined which has undecidable word problem. As noticed in [23], $\mathscr{M}$ can be seen as a 5-rule semi-Thue system which has undecidable accessibility problem. The system $\mathscr{M}$ encodes a system $\mathscr{C}$ defined in [9] such that for some word w the IWP for $(\mathscr{C}, w)$ is undecidable. From this w one can easily build a word w_0 such that the IPA for $(\mathscr{M}, w_0)$ is undecidable too.

3. Length-preserving semi-Thue systems with 9 rules and undecidable uniform termination problem

We give here a reduction of the *individual accessibility problem* for finite semi-Thue sytems to the *nonuniform termination problem* for length-preserving finite semi-Thue systems.[3]

The general idea of our reduction is the following.

Let $S_0 \subset X_0^* \times X_0^*$ be a finite semi-Thue system and w_0 a word over X_0. We build a semi-Thue system S_2 over a finite alphabet X_2 such that the IWP for (S_0, w_0) reduces to the IAP for (S_2, ε). Moreover $\xrightarrow{*}_{S_2}$ fulfills some kind of right-cancellation property (Lemma 4). We then associate to every $w \in X_2^*$, a length-preserving semi-Thue system $\mathscr{C}(w)$ such that,[4] $\mathscr{C}(w)$ is non-u-terminating if and only if $w \xrightarrow{*}_{S_2} \varepsilon$ (Lemma 7). Moreover $\mathrm{Card}(\mathscr{C}(w)) = \mathrm{Card}(S_0) + 4$.

Applying this reduction to the system $S_0 = \mathscr{M}$ recalled in Section 2.2, we prove the undecidability of the *uniform termination problem* for length-preserving finite semi-Thue systems having 9 rules.

We begin the constructions in details. Let $X_0 = \{a, b\}$, and S_0 be a finite semi-Thue system over X_0 and $w_0 \in X_0^*$, $X_1 = \{a, b, x, \bar{x}\}$, $S_1 = S_0 \cup \{xw_0\bar{x}, \varepsilon\}$ and

---

[3] We give here a *many-one* reduction of the IAP to the NUTP. Of course it implies that the IAP is *truth-table*-reducible (and a fortiori *Turing*-reducible) to the UTP, see [24, Chs. 7, 8, 9].

[4] This statement is proved for words w of a special kind only, see Lemma 7, but we omit here the technical details.

$\psi : \{a, b, x, \bar{x}\}^* \to \{a, b\}^*$ be the homomorphism defined by

$$\psi(a) = aabab^4, \quad \psi(b) = aab^2ab^3, \quad \psi(x) = aab^3ab^2, \quad \psi(\bar{x}) = aab^4ab.$$

We set $y = \psi(x)$, $\bar{y} = \psi(\bar{x})$. Furthermore, let $S_2 = \{(\psi(u), \psi(v)) \mid (u, v) \in S_1\}$.

Lemma 4. *Let $u \in \{a, b\}^*$, $v \in \{a, b\}^*$, $n \geqslant 1$. The following properties are equivalent:*

(1) $u(y\psi(v)\bar{y})^n \xrightarrow{*}_{S_2} u$,
(2) $y\psi(v)\bar{y} \xrightarrow{*}_{S_2} \varepsilon$,
(3) $\psi(v) \xrightarrow{*}_{S_2} \psi(w_0)$,
(4) $v \xrightarrow{*}_{S_0} w_0$.

Proof. Obviously, (4) $\Rightarrow$ (3) $\Rightarrow$ (2) $\Rightarrow$ (1). Let us suppose that (4) is false:

$$v \not\xrightarrow{*}_{S_0} w_0.$$

As ψ is overlap-free, we can conclude that (3) is false:

$$\psi(v) \not\xrightarrow{*}_{S_2} \psi(w_0).$$

Let us now show by induction on $m \geqslant 0$ the following property $P(m)$:

$\forall w \in \{a, b\}^*$, if $u(y\psi(v)\bar{y})^n \xrightarrow{m}_{S_2} w$ then $\exists u_0, v_1, \ldots, v_n \in \{a, b\}^*$ such that

$$w = u_0(y\psi(v_1)\bar{y}) \cdots (y\psi(v_i)\bar{y}) \cdots (y\psi(v_n)\bar{y}), \quad u \xrightarrow{*}_{S_2} u_0$$

$$\psi(v) \xrightarrow{*}_{S_2} \psi(v_i) \quad \text{for all } i \in [1, n].$$

If $m = 0$: this is clearly true.
If $m = m' + 1$:

$$u(y\psi(v)\bar{y})^n \xrightarrow{m'}_{S_2} w' \xrightarrow{1}_{S_2} w.$$

By induction hypothesis: $\exists u'_0, v'_1, \ldots, v'_n \in \{a, b\}^*$ such that

$$w' = u'_0(y\psi(v'_1)\bar{y}) \ldots (y\psi(v'_i)\bar{y}) \ldots (y\psi(v'_n)\bar{y}), \quad u \xrightarrow{*}_{S_2} u'_0$$

$$\psi(v) \xrightarrow{*}_{S_2} \psi(v'_i) \quad \text{for all } i \in [1, n]$$

- Every occurrence in w' of a word $\psi(\alpha)$, where $(\alpha, \beta) \in S_0$, is disjoint from every occurrence in w' of the words y, $\bar{y}$ (because $\psi(a)$ (resp. $\psi(b)$) has no overlap with $\psi(x)$ or with $\psi(\bar{x})$).
- As (3) is false, no $\psi(v'_i)$ is equal to $\psi(w_0)$.

Hence every redex (mod S_2) in w' must lie inside one of the given occurrences of u'_0, $\psi(v'_1), \ldots, \psi(v'_n)$; this establishes $P(m)$.

For every word $w \in \{a, b\}^*$, let us denote by $|w|_y$ the number of distinct occurrences of y as a factor of w. If w has the form given in $P(m)$, then $|w|_y \geqslant n$. Hence for $p \geqslant |u| + 1$,

$$u(y\psi(v)\bar{y})^p \not\xrightarrow{*}_{S_2} u.$$

But $u(y\psi(v)\bar{y})^n \xrightarrow{*}_{S_2} u$ would imply

$$u(y\psi(v)\bar{y})^p \xrightarrow{*}_{S_2} u, \quad \text{with } p = n(|u|+1).$$

Hence $u(y\psi(v)\bar{y})^n \not\xrightarrow{*}_{S_2} u$. □

By $\dot{-}$ we denote the substraction in $\mathbb{N}$:

$$p \dot{-} q = \begin{cases} 0 & \text{if } p < q, \\ p - q & \text{if } p \geqslant q. \end{cases}$$

To every $w \in \{a,b\}^*$ we associate the semi-Thue system $\mathscr{C}(w)$ over the alphabet $X = \{a,b,e,\#\}$ consisting of the set of rules:

(r1) $e^{|w|}\# \to w\#$,

(r2) $ea \to ae$,

(r3) $eb \to be$,

(R4) $e^{1+(|v| \dot{-} |u|)}u \to ve^{1+(|u| \dot{-} |v|)}$ (for every $(u,v) \in S_2$).

Let $\varphi : X^* \to \{a,b\}^*$ be the homomorphism defined by

$$\varphi(a) = a, \quad \varphi(b) = b, \quad \varphi(e) = \varphi(\#) = \varepsilon.$$

Lemma 5. *For every $w, w' \in X^*$*

$$(w \xrightarrow{*}_{\mathscr{C}(w) - \{r1\}} w') \Rightarrow (\varphi(w) \xrightarrow{*}_{S_2} \varphi(w')).$$

Proof. It suffices to check that, $\forall(\alpha,\beta) \in \mathscr{C}(w) - \{r1\}$, $\varphi(\alpha) \xrightarrow{n}_{S_2} \varphi(\beta)$, where $n \in \{0,1\}$. □

Lemma 6. *$\mathscr{C}(w) - \{r1\}$ is u-terminating.*

Proof. Let $\preceq$ be the lexicographic ordering on X^* induced by

$$e \prec a \prec b \prec \#.$$

Then, $\forall(\alpha,\beta) \in \mathscr{C}(w) - \{r1\}$,

$$|\tilde{\alpha}| = |\tilde{\beta}| \quad \text{and} \quad \tilde{\alpha} \preceq \tilde{\beta}.$$

Hence $\mathscr{C}(w) - \{r1\}$ is u-terminating. □

Lemma 7. *Let $v \in \{a,b\}^*$. $\mathscr{C}(y\psi(v)\bar{y})$ is not u-terminating $\Leftrightarrow$ $y\psi(v)\bar{y} \xrightarrow{*}_{S_2} \varepsilon$.*

Proof. ($\Leftarrow$) Let $w = y\psi(v)\bar{y}$. Let $w = v_0 \to_{S_2} v_1 \to_{S_2} \cdots v_i \to_{S_2} v_{i+1} \cdots \to_{S_2} v_n = \varepsilon$ be a derivation from w to ε. Let $l = \max\{1 + (|v| \dot{-} |u|) \mid (u,v) \in S_2\}$ and $p = ln$. One can prove by induction on i that:

$$\forall i \in [0,n],\ \exists q_i \geqslant 0, \quad e^p w\# \xrightarrow{*}_{\mathscr{C}(w)} e^{p-li} v_i e^{q_i}\#.$$

As $\mathscr{C}(w)$ preserves the length $q_n = p + |w|$, hence

$$v_n e^{q_n}\# = e^{p+|w|}\# \to^1_{\{r1\}} e^p w\#.$$

Hence $e^p w\# \to^+_{\mathscr{C}(w)} e^p w\#$.

($\Rightarrow$) For every derivation $D = (v_i)_{i\in I}$ we note $\|D\| = |v_0|$ (we call $\|D\|$ the width of D and notice that $\forall i\in I$, $\|D\| = |v_i|$ because $\mathscr{C}(w)$ is length-preserving). Let $D = (v_i)_{i\in I}$ be an infinite derivation of minimal width. The form of the rules implies then:

$$\forall i\in I,\quad v_i\in\{a,b,e\}^*\#.$$

By Lemma 6, infinitely many steps of D use the rule r1.

Hence there exists an infinite sequence $i_1 < i_2 < \cdots < i_j < \cdots$ such that, for every $j \geqslant 1$,

$$v_{i_j} = u_j w\# \to^*_{\mathscr{C}(w)-\{r1\}} u_{j+1} e^{|w|}\# = v_{i_{j+1}-1}.$$

By Lemma 5, for every $j \geqslant 1$,

$$\varphi(u_j)w \xrightarrow{*}_{S_2} \varphi(u_{j+1}).$$

As the words $\{\varphi(u_j) \mid 1 \leqslant j\}$ have bounded length, there exists $1 \leqslant j < j'$ such that

$$\varphi(u_j) = \varphi(u_{j'}),$$

which implies that

$$\varphi(u_j)w^{j'-j} \xrightarrow{*}_{S_2} \varphi(u_j).$$

But $w = y\psi(v)\bar{y}$, hence by Lemma 4, points (1) and (2),

$$w \xrightarrow{*}_{S_2} \varepsilon. \qquad \square$$

Theorem 8. *If there exists some semi-Thue system S_0 with n rules and some word w_0 such that the individual word problem for (S_0, w_0) is undecidable, then the uniform termination problem is undecidable for length-preserving semi-Thue systems with $n + 4$ rules.*

Proof. By Lemma 4 "$v \xrightarrow{*}_{S_0} w_0$?" reduces to "$y\psi(v)\bar{y} \xrightarrow{*}_{S_2} \varepsilon$?" which by Lemma 7 reduces to "is $\mathscr{C}(y\psi(v)\bar{y})$ u-terminating?". Hence this last problem is undecidable. One can check that $\mathscr{C}(w)$ has $\mathrm{Card}(S_0) + 4$ rules. $\square$

Theorem 9. *The uniform termination problem is undecidable for length-preserving semi-Thue systems having 9 rules.*

Proof. Let us take $S_0 = \mathscr{M}$ and w_0 be a word such that the IWP for (S_0, w_0) is undecidable (see Section 2.2). We recall that $\mathscr{M}$ has 5 rules. The result follows from Theorem 8. $\square$

4. Semi-Thue systems with 10 rules and undecidable termination problem

In this section we build explicitly a semi-Thue system $\mathscr{T}$ which has 9 rules and such that the termination problem of any system S on any word t can be encoded as the termination problem of $\mathscr{T} \cup \{u_S \to v_S\}$ on a word $w_{S,t}$ (where u_S, v_S, $w_{S,t}$ can be computed from S and t).

Let us consider the alphabet $X = \{a, b, c, e\}$ and the system $\mathscr{S}$ on X composed of the following rules:

$$\begin{array}{lll} \mathrm{r1}: cae \to eaa, & \mathrm{r5}: eaa \to ae, & \mathrm{r9}: ac \to ca, \\ \mathrm{r2}: cbe \to eab, & \mathrm{r6}: eab \to be, & \mathrm{r10}: bc \to cb, \\ \mathrm{r3}: cae \to eba, & \mathrm{r7}: eba \to ae, & \mathrm{r11}: ec \to ce, \\ \mathrm{r4}: cbe \to ebb, & \mathrm{r8}: ebb \to be, & \mathrm{r12}: cce \to ec, \\ & & \mathrm{r13}: ccac \to acc. \end{array}$$

4.1. Some properties of $\mathscr{S}$

Lemma 10. *$\mathscr{S}$ is uniformly terminating.*

Proof. Let $\preceq_{\mathrm{lex}}$ be the lexicographic ordering on X^* induced by: $c \prec b \prec a \prec e$. Let $\preceq$ be the ordering on X^* defined by $f \preceq g$ iff ($|f| < |g|$ or ($|f| = |g|$ and $f \preceq_{\mathrm{lex}} g$). Let $\lambda: X^* \to \mathbb{N} \times X^*$ defined by $\lambda(f) = (|f|_c, f)$. Let $\sqsubseteq$ be the lexicographic ordering on $\mathbb{N} \times X^*$ deduced from the usual ordering on $\mathbb{N}$ and the ordering $\preceq$ on X^*: for every $(n,u), (n',u') \in \mathbb{N} \times X^*$,

$$(n,u) \sqsubseteq (n',u') \Leftrightarrow n \leqslant n' \text{ or } (n = n' \text{ and } u \preceq u').$$

If $u \to_{\mathscr{S}} v$ then $\lambda(u) \sqsupset \lambda(v)$. Hence $\mathscr{S}$ is uniformly terminating. □

From now on we consider some fixed integers $k \geqslant 0$, $l \geqslant 0$.

Let $\Pi: \{a,b,c,e\}^* \to \{a,b,e\}^*$ be the homomorphism defined by: $\forall x \in \{a,b,e\}$, $\Pi(x) = x$ and $\Pi(c) = \varepsilon$.

Let $\Omega \subset \{a,b,c,e\}^*$ defined by: $\Omega = \Pi^{-1}([(a+b)^* e^k]_{\overset{*}{\to}_{\mathscr{U}}})$ ($\mathscr{U}$ has been defined in Section 2).[5]

We define a function $H: \Omega \to \{a,b\}^*$ by

$$\forall w \in \Omega, \quad H(w) = \rho(\Pi(w)) e^{-k}$$

---

[5] One can notice that, since $\mathscr{U}$ is confluent and u-terminating, and since $\Omega_1 := (a+b)^* e^k$ consists of $\mathscr{U}$-irreducible words only, a word w belongs to Ω if and only if $\Pi(w) \overset{*}{\to}_{\mathscr{U}} u$ for some $u \in \Omega_1$. Thus, Ω really consists of all ancestors of the words from Ω_1 modulo $\mathscr{U}$ interspersed with an arbitrary number of occurrences of the letter c.

(in other words: $H(w)$ is the unique word such that $\Pi(w) \xrightarrow{*}_{\mathscr{U}} H(w)e^k$). We consider the integers: $P = (l+2)8^k$, $Q = (l+1)(2^k - 1) + 2$.

Lemma 11. *$\mathscr{S}$ preserves the set Ω and the function H i.e. $\forall w \in \Omega$, $\forall w' \in \{a,b,c,e\}^*$, if $w \xrightarrow{*}_{\mathscr{S}} w'$ then*

(1) $w' \in \Omega$,

(2) $H(w) = H(w')$.

Proof. $\Pi(\mathscr{S})$ and $\mathscr{U}$ are generating the same congruence on $\{a,b,e\}^*$, hence $\mathscr{S}$ preserves Ω and also the function: $w \mapsto \rho(\Pi(w))$, hence it preserves H. □

Lemma 12. *Let $p \geqslant 0$, $q = p.2^k$ and $a_1, a_2, \ldots, a_p \in \{a,b\}$. Then there exist $r \geqslant 0$ and $b_1, b_2, \ldots, b_q \in \{a,b\}$ such that*

$$a_1 c^P a_2 c^P \cdots a_p c^P e^k \xrightarrow{*}_{\mathscr{S}} c^r e^k b_1 c^Q b_2 c^Q \cdots b_q c^Q.$$

Proof. For every $n \geqslant 0$ we note $P(n) = 4^n Q + (4^n - 1)/3$. $\forall n \geqslant 1$, $\forall a_1, b_1 \in \{a,b\}$, $a_1 c^{P(n)} e \xrightarrow{*}_{\mathscr{S}} e b_1 c^{(P(n)-1)/4} a_1 c^{(P(n)-1)/4} = e b_1 c^{P(n-1)} a_1 c^{P(n-1)}$. By induction on n and p we obtain:

$\forall a_1, a_2, \ldots, a_p \in \{a,b\}, \exists b_1, b_2, \ldots, b_q \in \{a,b\}$ (with $q = p2^n$) such that

$$a_1 c^{P(n)} a_2 c^{P(n)} \cdots a_p c^{P(n)} e^k \xrightarrow{*}_{\mathscr{S}} e^k b_1 c^{P(n-k)} b_2 c^{P(n-k)} \cdots b_q c^{P(n-k)}. \tag{6}$$

Let us take $n = k$ in (1), define $r = p(P - P(k))$, notice that $Q = P(0)$. We obtain

$$a_1 c^P a_2 c^P \ldots a_p c^P e^k \xrightarrow{*}_{\mathscr{S}} c^r a_1 c^{P(k)} a_2 c^{P(k)} \ldots a_b c^{P(k)} e^k \xrightarrow{*}_{\mathscr{S}} c^r e^k b_1 c^Q b_2 c^Q \ldots b_q c^Q. \qquad \square$$

Lemma 13. *Let $w = ue^k v, \alpha, A, \beta \in \{a,b\}^*$, $x \in \{a,b\}$ such that*

(1) $u \in c^*((a+b)c^P)^*$,

(2) $v \in ((a+b)c^Q)^*$,

(3) $H(w) = \alpha A x \beta$,

(4) $|\Pi(u)| < |\alpha A x|$,

then there exists u_1, u_2, v_1 such that

(5) $ue^k v \xrightarrow{*}_{\mathscr{S}} u_1 e^k c^Q v_1$,

(6) $u_2 \in c^* \{a,b\}^*$,

(7) $u_1 = u u_2$,

(8) $v_1 \in ((a+b)c^Q)^*$,

(9) $\Pi(u_1) = \alpha A x$.

Proof. Let v_1, v_2 be the factors of v such that $|\Pi(v_2)| = (|\alpha A x| - |\Pi(u)|)2^k$ and $v = v_2 c^Q v_1$. Let $u_3 = H(e^k v_2)$. Then $\exists r \geqslant 0$, $e^k v_2 \xrightarrow{*}_{\mathscr{S}} c^r u_3 e^k$ (it suffices to use rules r5–r11).

Let us take $u_2 = c^r u_3$ and $u_1 = u u_2$. We have

$$ue^k v = ue^k v_2 c^Q v_1 \xrightarrow{*}_{\mathscr{S}} u u_2 e^k c^Q v_1 = u_1 e^k c^Q v_1$$

(see Fig. 1). □

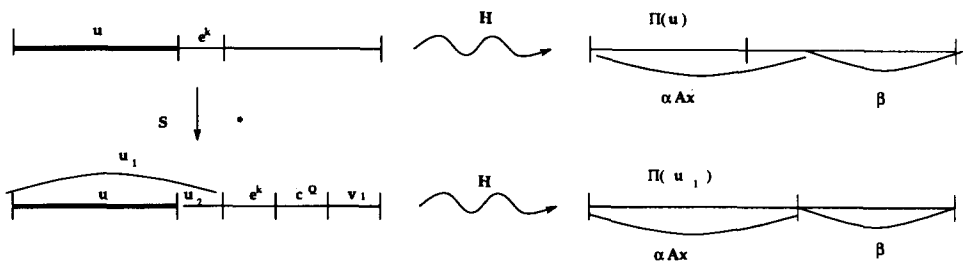

Fig. 1. Lemma 13

Lemma 14. *Let* $A\in\{a,b\}^*$, $x\in\{a,b\}$, $W\in\{a,b\}^*$ *such that* $|A|=l$ *and* $Axe^k \overset{*}{\leftarrow}_{\mathscr{R}} e^kW$. *Then* $Axe^kc^Q \overset{*}{\rightarrow}_{\mathscr{S}} e^kWc^2$.

Proof. Let $K=(l+1)+2(l+1)+\cdots+2^{k-1}(l+1)=(l+1)(2^k-1)$. One can check that

$$Axe^kc^K \overset{*}{\rightarrow}_{\mathscr{S}} e^kW.$$

As $Q=K+2$, the lemma follows. □

4.2. Universality of $\mathscr{S}$.

Let us consider two integers $h\geqslant 2$, $m\geqslant 0$. Let us define: $D_h=\{aab^iab^{h+1-i}|\ 1\leqslant i\leqslant h\}$ (this code is introduced in [21]). $\Omega(h)=\Pi^{-1}([D_h^*ae^k]_{\overset{*}{\leftarrow}_{\mathscr{R}}})$ (we can notice that $\Omega(h)\subset\Omega$).

Let $T\subset\{a,b\}^*\times\{a,b\}^*$ be a semi-Thue system such that $T=\{(A_j,B_j)|\ 1\leqslant j\leqslant 2^k\}$ where for every j,

$$A_j,B_j\in D_h^*,\qquad |A_j|=l,\qquad |B_j|=m,$$

$$A_j=a_{1,j}a_{2,j}\ldots a_{l,j}\quad (a_{i,j}\in\{a,b\}),$$

$$B_j=b_{1,j}b_{2,j}\ldots b_{m,j}\quad (b_{i,j}\in\{a,b\}).$$

Let us then define (as in [21]):

$$L=a_{1,1}a_{1,2}\ldots a_{1,2^k}a_{2,1}a_{2,2}\ldots a_{2,2^k}\ldots a_{l,1}a_{l,2}\ldots a_{l,2^k},$$

$$M=b_{1,1}b_{1,2}\ldots b_{1,2^k}b_{2,1}b_{2,2}\ldots b_{2,2^k}\ldots b_{m,1}b_{m,2}\ldots b_{m,2^k}.$$

We consider the integer $R=(l+m+2)P$ and the one-rule semi-Thue system: $\mathscr{R}=\{(Lc,Mc^R)\}$.

Lemma 15. *Let* $w\in\Omega(h)$ *and let* $w \overset{*}{\rightarrow}_{\mathscr{S}\cup\mathscr{R}} w'$ *be a derivation with exactly* n *steps in* $\rightarrow_{\mathscr{R}}$. *Then* $H(w) \overset{n}{\rightarrow}_T H(w')$.

Proof. Lemma 11 remains true with $\Omega(h)$ in place of Ω. Since L admits the prefix $a^{2^{k+1}}$, every occurrence of L in a word $w \in \Omega(h)$ must be on the right of the rightmost occurrence of e. The "Folding lemma" (see Section 2) then shows that

$$w \xrightarrow{1}_{\mathscr{R}} w' \Rightarrow H(w) \xrightarrow{1}_{T} H(w'). \qquad \square$$

Lemma 16. *Let* $w = ue^k v$, $j \in [1, 2^k]$, $\alpha, \beta \in \{a, b\}^*$ *such that*

(1) $u \in c^*((a+b)c^P)^*$,
(2) $v \in ((a+b)c^Q)^*$,
(3) $H(w) = \alpha A_j a \beta$,
(4) $|\alpha A_j a| \leqslant |\Pi(u)| + l - 1$

then there exists u_1, v_1 *such that*

(5) $ue^k v \to^{+}_{\mathscr{P} \cup \mathscr{R}} u_1 e^k v_1$,
(6) $u_1 \in c^*((a+b)c^P)^*$,
(7) $v_1 \in ((a+b)c^Q)^*$,
(8) $\Pi(u_1) = \alpha B_j a$,
(9) $H(e^k v_1) = \beta$.

This lemma is a kind of converse of Lemma 15. It shows that, if w, w_1 are words in a suitable rational subset of $\Omega(h)$, namely the set

$$\Omega(h) \cap c^*((a+b)c^P)^* e^k ((a+b)c^Q)^*$$

then, every rewriting step $H(w) \to_T h_1$ can be lifted to a derivation $w \to^{+}_{\mathscr{P} \cup \mathscr{R}} w_1$, with $H(w_1) = h_1$. The technical hypothesis (4) expresses the fact that the rewriting step $H(w) \to_T h_1$ under scrutiny does not use a redex which is "*too far on the right*" of the factor $\Pi(u)$ in $H(w) = \Pi(u)H(e^k v)$. This technical hypothesis will be satisfied by each step of a *right-to-left* derivation (mod T) (see proof of Lemma 17).

Proof. We distinguish two cases, according to the location of the redex A_j relative to $\Pi(u)$. Either $|\alpha A_j a| \leqslant |\Pi(u)|$ (case 1), which is the simplest case to handle. Or $|\alpha A_j a| > |\Pi(u)|$ (case 2). This case is more delicate to handle. We shall use the technical hypothesis (4), and the fact that the integer R is "sufficiently large" to ensure that we can keep the word w_1 in the rational set $c^*((a+b)c^P)^* e^k ((a+b)c^Q)^*$.

Case 1: $|\alpha A_j a| \leqslant |\Pi(u)|$. Let $u' \in c^*((a+b)c^P)^*$, $u'' \in ((a+b)c^P)^* \mid u = u'u''$ and $\Pi(u') = |\alpha A_j a|$. Applying Lemma 12 to $u'' = a_1 c^P a_2 c^P \cdots a_p c^P$ we get some word $v'' \in ((a+b)c^Q)^*$ and some $r \geqslant 0$ such that

$$u'' e^k \xrightarrow{*}_{\mathscr{P}} c^r e^k v''.$$

Hence

$$ue^k v \xrightarrow{*}_{\mathscr{P}} c^r u' e^k v'' v \quad \text{with } \Pi(u') = \alpha A_j a. \tag{7}$$

Let $u' = u'_1 u'_2$ where $\Pi(u'_1) = \alpha$, $\Pi(u'_2) = A_j a$, $u'_2 \in ((a+b)c^P)^*$.

$$u'_2 e^k = a_{1,j} c^P a_{2,j} c^P \cdots a_{l,j} c^P a c^P e^k \xrightarrow{*}_{\mathscr{P}} c^{r_1} a_{1,j} a_{2,j} \cdots a_{l,j} a e^k c^Q \tag{8}$$

(where $r_1 = (l+1)P - 2^k Q$).

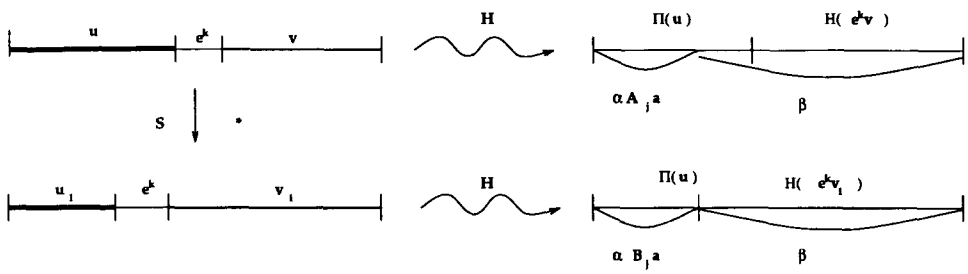

Fig. 2. Lemma 16, case 1

By the Folding lemma

$$A_j a e^k \overset{*}{\leftarrow}_{\mathscr{U}} e^k u_3 L v_3 a \tag{9}$$

(for some u_3, $v_3 \in \{a,b\}^*$, with $|u_3| \equiv -j \pmod{2^k}$, $|u_3| + |v_3| = 2^k - 1$). By (9) and Lemma 14

$$A_j a e^k c^Q \overset{*}{\rightarrow}_{\mathscr{S}} e^k u_3 L v_3 a c^2, \tag{10}$$

$$e^k u_3 L v_3 a c^2 \overset{*}{\rightarrow}_{\mathscr{S}} e^k u_3 (Lc) v_3 a c \overset{1}{\rightarrow}_{\mathscr{R}} e^k u_3 M c^R v_3 a c. \tag{11}$$

By the Folding Lemma:

$$e^k u_3 M c^R v_3 a c \overset{*}{\rightarrow}_{\mathscr{S}} B_j c^R a c e^k. \tag{12}$$

(Using rules r5–r8, the four analogous derivations

$$e x c^R y \overset{*}{\rightarrow}_{\mathscr{S}} c^R y e \quad (x, y \in \{a,b\})$$

and also rule r11.)

Let $r_2 = R - mP - 2(P-1)$ and $u_2'' \in ((a+b)c^P)^*$ such that $\Pi(u_2'') = B_j a$:

$$B_j c^R a c \overset{*}{\rightarrow}_{\mathscr{S}} c^{r_2} u_2'' \tag{13}$$

(using r9, r10, r13). By (7), (8), (10)–(13)

$$u e^k v \rightarrow^{+}_{\mathscr{S} \cup \mathscr{R}} c^r u_1' c^{r_1 + r_2} u_2'' e^k v'' v. \tag{14}$$

Using rules r9, r10,

$$c^r u_1' c^{r_1 + r_2} u_2'' e^k v'' v \overset{*}{\rightarrow}_{\mathscr{S}} c^{r + r_1 + r_2} u_1' u_2'' e^k v'' v. \tag{15}$$

The equation (7) and the invariance of H by $\mathscr{S}$ imply that $H(e^k v'' v) = \beta$, hence $u_1 = c^{r + r_1 + r_2} u_1' u_2''$ and $v_1 = v'' v$ are the required words.

Case 2: $|\Pi(u)| + 1 \leqslant |\alpha A_j a| \leqslant |\Pi(u)| + l - 1$. By Lemma 13 there exist u', u_4, v_4 such that

$$u e^k v \overset{*}{\rightarrow}_{\mathscr{S}} u' e^k c^Q v_4 \tag{16}$$

(with $u_4 \in c^* \{a,b\}^*$, $u' = u u_4$, $v_4 \in ((a+b)c^Q)^*$, $\Pi(u') = \alpha A_j a$). Moreover here,

$$|\Pi(u_4)| \leqslant l - 1. \tag{17}$$

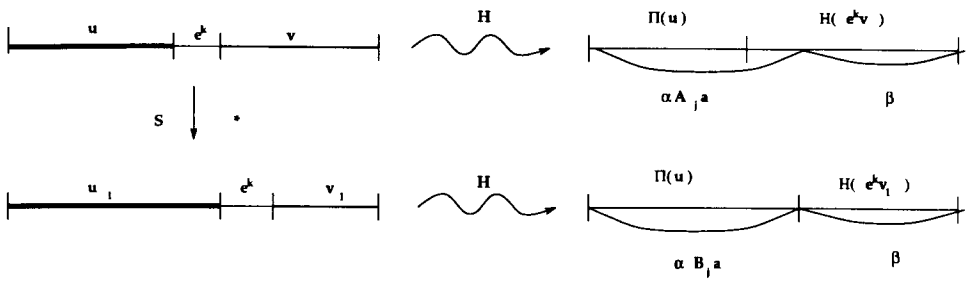

Fig. 3. Lemma 16, case 2

By arguments similar to those used in (8)–(12) we obtain: $\exists u_5 \in ((a+b)c^P)^*$, $\exists u_6 \in (a+b)^*$, $\exists r_3 \geqslant 0$ such that

$$u' e^k c^Q v_4 \to^+_{\mathscr{S} \cup \mathscr{R}} c^{r_3} u_5 u_6 c^R a c e^k v_4, \tag{18}$$

$\Pi(u_5 u_6 a) = \alpha B_j a$, $|u_6| \leqslant m + l - 1$. As $R \geqslant (m+l-1)P + 2(P-1)$, $\exists r_4 \geqslant 0$ such that

$$u_6 c^R a c \to^*_{\mathscr{S}} c^{r_4} u'_6 a c^P \quad \text{with } u'_6 \in ((a+b)c^P)^*. \tag{19}$$

By (16)–(19)

$$u e^k v \to^*_{\mathscr{S} \cup \mathscr{R}} c^{r_3 + r_4} u_5 u'_6 a c^P e^k v_4. \tag{20}$$

Hence $u_1 = c^{r_3 + r_4} u_5 u'_6 a$ and $v_{12} = v_4$ are the required words. □

Let $\phi : \{a,b\}^* \to \{a,b,c\}^*$ be the homomorphism defined by $\phi(a) = ac^P$, $\phi(b) = bc^P$.

Lemma 17. *Let $u \in D_h^*$. T has an infinite derivation starting on u iff $\mathscr{S} \cup \mathscr{R}$ has an infinite derivation starting on $\phi(ua)e^k$.*

Proof. If T has an infinite derivation starting on u, by Lemma 1, T has an infinite rl-derivation starting on u, hence an infinite rl-derivation starting on ua. As D_h is overlap-free, every redex A_j in this derivation has a right-context beginning with a. As the derivation is rl, if $w = \alpha B_i a \beta = \alpha' A_j a \beta'$ (where B_i is the result of a derivation-step and A_j the redex of the next step), then $|\alpha' A_j a| \leqslant |\alpha B_i a| + l - 1$. Lemma 16 shows that each step can be "lifted" to a derivation (mod $\mathscr{S} \cup \mathscr{R}$) with nonnull, finite length. Hence $\mathscr{S} \cup \mathscr{R}$ has an infinite derivation starting on $\phi(ua)e^k$.

The converse statement comes from Lemmas 9 and 14 and the fact that, in a T-derivation starting from ua, no redex, can use the rightmost letter a (because $T \subset D_h^* \times D_h^*$). □

4.3. The system $\mathscr{T}$

Let $\tau : \{a,b,c,e\}^* \to \{a,b,c\}^*$ be the injective homomorphism defined by $\tau(a) = a$, $\tau(b) = ab$, $\tau(c) = c$, $\tau(e) = bb$. Let

$$\mathscr{T} = \{\tau(\mathrm{r1}), \tau(\mathrm{r3}), \tau(\mathrm{r5}), \tau(\mathrm{r7}), \tau(\mathrm{r9}), \tau(\mathrm{r10}), \tau(\mathrm{r11}), \tau(\mathrm{r12}), \tau(\mathrm{r13})\}.$$

Hence $\mathcal{T}$ consists of the following rules:

$$\begin{array}{lll} r'1: cabb \to bbaa, & r'5: bbaa \to abb, & r'9: ac \to ca, \\ & & r'10: abc \to cab, \\ r'3: cabb \to bbaba, & r7: bbaba \to abb, & r'11: bbc \to cbb, \\ & & r'12: ccbb \to bbc, \\ & & r'13: ccac \to acc. \end{array}$$

Theorem 18. *The system $\mathcal{T}$ is* "universal with respect to termination" *in the following sense: the termination problem of any system S on any word t can be encoded as the termination problem of $\mathcal{T} \cup \{u_S \to v_S\}$ on a word $w_{S,t}$ (where the words u_S, v_S, $w_{S,t}$ are computable functions of S and t).*

Proof. Let X be a finite alphabet and S a finite semi-Thue system on X.

$$S: u_i \to v_i \quad (1 \leqslant i \leqslant n).$$

Let us consider

$$l = \max\{|u_i| \mid 1 \leqslant i \leqslant n\}, \qquad m = \max\{|v_i| \mid 1 \leqslant i \leqslant n\}, \qquad Y = X \cup \{x_1\}.$$

We then define a new system S_1 over the alphabet Y by

$$S_1: u_i\beta \to v_i\beta x_1^{l+m-|v_i\beta|} \quad (\forall i \in [1,n],\ \forall \beta \in Y^{l-|u_i|}),$$

$$x_1\gamma \to \gamma x_1^{m+1} \quad (\forall \gamma \in Y^{l-1}).$$

This system S_1 is homogeneous and such that, for every $u \in X^*$,

$$S \text{ has an infinite derivation starting on } u$$
$$\Leftrightarrow\ S_1 \text{ has an infinite derivation starting on } ux_1^l. \tag{21}$$

Let us note $h = |Y|$ and $Y = \{y_1, y_2, \ldots, y_h\}$. We define an homomorphism $\psi: Y^* \to \{a,b\}^*$ by

$$\forall i \in [1,h], \quad \psi(y_i) = aab^i ab^{h+1-i}.$$

Let us define

$$T = \{(\psi(u), \psi(v)) \mid (u,v) \in S_1\}, \quad k = \lceil \log_2(\mathrm{Card}(T)) \rceil.$$

Let L, M, R be associated to T as defined in Section 4.2 and let

$$u_s = \tau(Lc), \quad v_S = \tau(Mc^R).$$

For every $u \in X^*$, the following properties are equivalent:

(0) S has an infinite derivation starting on u,
(1) S_1 has an infinite derivation starting on ux_1^l,
(2) T has an infinite derivation starting on $\psi(ux_1^l)$,

(3) $\mathscr{S} \cup \mathscr{R}$ has a infinite derivation starting on $\phi(\psi(ux_1^l)a)e^k$,
(4) $\tau(\mathscr{S} \cup \mathscr{R})$ has an infinite derivation starting on $\tau(\phi(\psi(ux_1^l)a)e^k)$.
(5) $\mathscr{T} \cup \{(u_S, v_s)\}$ has an infinite derivation starting on $\tau(\phi(\psi(ux_1^l)a)e^k)$.

(0) $\Leftrightarrow$ (1) holds by (21),
(1) $\Leftrightarrow$ (2) holds because ψ is overlap-free,
(2) $\Leftrightarrow$ (3) holds by lemma 17,
(3) $\Leftrightarrow$ (4) holds because $\mathscr{S} \cup \mathscr{R}$ fulfills the hypothesis of the Lifting Lemma,
(4) $\Leftrightarrow$ (5) holds because

$$\rightarrow^+_{\tau(\mathscr{S} \cup \mathscr{R})} = \rightarrow^+_{\mathscr{T} \cup \{(u_S, v_S)\}}.$$

The three words $\tau(\phi(\psi(ux_1^l)a)e^k), u_S, v_S$ are clearly computable from S, u. The equivalence between (0) and (5) then proves Theorem 18. □

Theorem 19. *There exists two words $u_0, v_0 \in \{a, b, c\}^*$ such that the termination problem for $\mathscr{T} \cup \{(u_0, v_0)\}$ is undecidable.*

Proof. Let S_0 be a finite semi-Thue system on X such that the termination problem for S_0 is undecidable (such an S_0 exists, see [19]) and let $(u_0, v_0) = (u_{S_0}, v_{S_0})$. By Theorem 18, the termination problem for $\mathscr{T} \cup \{(u_0, v_0)\}$ is undecidable. □

As $\mathscr{T}$ posesses 9 rules, the above theorem shows the existence of a fixed semi-Thue system with 10 rules and undecidable termination problem.

Acknowledgements

I thank Y. Métivier for fruitful discussions about termination problems. Theorems 9 and 19, as well as their proofs are essentially the ones which were presented at RTA93 (though theorem 9 is slightly improved here). I thank Y. Matijasevich for many fruitful comments on a previous version of the paper. Subsequent improvements of the bounds by Y. Matijasevich and the author will be presented in a forthcoming work. I am indebted to one of the anonymous referees for reading thoroughly a previous version of the manuscript and correcting several mistakes.

References

[1] M. Billaud, P. Lafon, Y. Métivier and E. Sopena, Graph rewriting systems with priorities, in: *Proc. WG 89*, Lecture Notes in Computer Science, Vol. 411 (Springer, Berlin, 1989) 94–106.
[2] R.V. Book, Homogeneous Thue systems and the Church–Rosser property, *Discrete Math.* (1984) 137–145.
[3] R.V. Book and F. Otto, Cancellation rules and extended word problems, *Inform. Process. Lett.* **20** (1985) 5–11.

[4] R.V. Book and F. Otto, *String Rewriting Systems. Texts and monographs in Computer Science* (Springer, Berlin, 1993).
[5] W.W. Boone, D. Collins and Y.V. Matiyasevich, Embedding into semi-groups with only a few defining relations, in: Proc. 2nd Scandinavian Logic Symp. Studies in Logic and the foundations of Mathematics (North-Holland, Amsterdam, 1971) 27–40.
[6] V.V. Borisov, Simple examples of groups with unsolvabe word problem, *Mat. Zametki* **6**; an English translation appears in: Math. Notes **6** (1969) 521–532.
[7] F. Bossut, M. Dauchet and B. Warin, Automata and rational expressions on planar graphs, Lecture Notes in Computer Science (Springer, Berlin, 1988).
[8] A.C. Caron, Linear bounded automata and rewrite systems: influence of initial configurations on decision properties, in: *Proc. CAAP* 91, Lecture Notes in Computer Science (Springer, Berlin, 1991).
[9] G.S. Ceitin, An associative calculus with an insoluble problem of equivalence, *Trudy Mat. Inst. Steklov.* **52** (1952) 172–189.
[10] D.J. Collins, Word and conjugacy problems in groups with only a few defining relations, *Z. Math. Logik Grundlag. Math.* **15**(4) (1969) 305–323.
[11] D.J. Collins, A. simple presentation of a group with unsolvable word problem, *Illinois J. Math.* **30**(2) (1986) 230–234.
[12] M. Dauchet, Simulation of Turing machines by a left-linear rewrite rule, in: *Proc. RTA* 89, Lecture Notes in Computer Science, Vol. 355 (Springer, Berlin, 189) 109–120.
[13] M. Davis, *Computability and Unsolvability* (McGraw-Hill, New York, 1958).
[14] N. Dershowitz, Termination of rewriting, *J. Symbolic Comput.* **3** (1987) 69–116.
[15] N. Dershowitz and J.P. Jouannaud, Rewrite sysrtems, in: *Handbook of Theoretical Computer Science, Vol. B*, Ch. 2 (Elsevier, Amsterdam, 1991) 243–320.
[16] P.K. Hooper, The undecidability of the Turing machine immortality problem, *J. Symbolic Logic*, **31**(2) (1966) 219–234.
[17] G. Huet, Confluent reductions: abstract properties and applications to term rewriting systems, *J. ACM* **27**(4) (1980) 797–821.
[18] G. Huet and D. Lankford, *On the uniform halting problem for term rewriting systems*, Rapport Laboria, 1978.
[19] M. Jantzen, *Confluent String Rewriting* (*EATCS monograph*) (Springer, Berlin, 1988).
[20] D. König, *Theorie der Endlichen und Unendlichen Graphen* (Chelsea, New York, 1950).
[21] J.V. Matiyasevich, Simple examples of undecidable associative calculi, *Soviet Math. Dokl.* (1967) 555–557.
[22] Y. Matiyasevich, Word problem for Thue systems with a few relations, in: *Actes de l'école de printemps d'informatique théorique, Font-Romeu* 93 (Springer, Berlin, 1994).
[23] J.J. Pansiot, A note on Post's correspondence problem, *IPL* **12** (1981) 233–233.
[24] H.J. Rogers, *Theory of Recursive Functions and Effective Computability* (McGraw-Hill, 1967).